World Academy of Science, Engineering and Technology
International Journal of Computer and Systems Engineering
Vol:16, No:10, 2022

Using Satellite Images Datasets for Road Intersection Detection in Route Planning

Fatma El-zahraa El-taher, Ayman Taha, Jane Courtney, Susan Mckeever

Abstract—Understanding road networks plays an important role in navigation applications such as self-driving vehicles and route planning for individual journeys. Intersections of roads are essential components of road networks. Understanding the features of an intersection, from a simple T-junction to larger multi-road junctions is critical to decisions such as crossing roads or selecting safest routes. The identification and profiling of intersections from satellite images is a challenging task. While deep learning approaches offer state-of-the-art in image classification and detection, the availability of training datasets is a bottleneck in this approach. In this paper, a labelled satellite image dataset for the intersection recognition problem is presented. It consists of 14,692 satellite images of Washington DC, USA. To support other users of the dataset, an automated download and labelling script is provided for dataset replication. The challenges of construction and fine-grained feature labelling of a satellite image dataset are examined, including the issue of how to address features that are spread across multiple images. Finally, the accuracy of detection of intersections in satellite images is evaluated.

Keywords—Satellite images, remote sensing images, data acquisition, autonomous vehicles, robot navigation, route planning, road intersections.

I. Introduction

ROAD networks are divided into two main components; roads segments and junctions. Junctions, termed ”intersections” in the navigation literature, are critical areas in a road network where vehicles, robots or people take active decisions, such as crossing roads and changing directions.

Intersections include three features: location, degree (number of branches) and shape. The location is defined by the longitude and latitude of an intersection's centre. The degree determines the number of segments of roads that meet together to create the intersection. The intersection's shape is an angular arrangement of its branches, such as a T or Y junction [1], [2].

Predefining intersection features increases the flexibility of route planning so users can include or exclude intersections according to their preferences. For vehicles, signal intersections increase journey time and fuel consumption [3], [4]. For people, especially the visually impaired, the inclusion of simple intersections and the exclusion of complex intersections increases the safety of their journey [5], [6]. Identification of intersections is a key task to extract road networks from satellite images, especially in a complex urban area [7], [8].

Fatma El-zahraa El-taher and Susan Mckeever are with the School of Computer Science, Technological University Dublin, Dublin, Ireland (e-mail: fatma.e.eltaher@myTUDublin.ie).

Ayman Taha is with the School of Computer Science, Technological University Dublin, Dublin, Ireland and Faculty of Computers and Artificial Intelligence, Cairo University, Egypt.

Jane Courtney is with the School of Electrical Electronic Engineering, Technological University Dublin, Dublin, Ireland.

Satellite images are a promising source of data for identifying intersections' features, with three particular advantages. First, they capture different intersection structures and types. Secondly, they provide global coverage (e.g. Google Maps images), allowing for widely applicable detection approaches and models. Thirdly, they can be used in an offline process to extract map information.

This article aims to test the ability of deep learning models to identify the existence of intersections in satellite images, paving the way to enhanced route planning and road network extraction. Fit-for-purpose datasets are, in general, a bottleneck for machine learning approaches, particularly for the more data-reliant deep learning approaches. Benchmark datasets also serve to allow evaluation of the progress of a research community in a particular problem. Previously, used satellite image datasets have been very limited in terms of the number of images [8], [9], [10], which has been a limitation for progress in this area. In this paper, the largest satellite image dataset to date, for use in the intersection identification problem, is presented. In addition, the problem of objects' partial coverage in a satellite image is addressed. The dataset is used in the creation of several deep learning models, and the accuracy of each in detecting intersections is determined.

The structure of this paper is as follows. Section II discusses related work. Construction of datasets, including capture, annotation, and availability is explained in Section III. In Section IV, details of experiments are defined, while results are displayed and examined in Sections V and VI. Finally, the main points are highlighted in Section VII.

II. Related Work

Research work in detection of road features and intersections can be grouped based on the data source used, the type of task or the purpose of the intersection detection and the approaches used.

In terms of data, a variety of types have been used for intersection detection to date, including images [11], Open Street Map (OSM) data [2], videos [12], [13], [14], LiDAR sensors [15], and vehicle trajectories [16], [17], [18]. LiDAR sensors are very expensive, limiting their scope for blanket coverage. Vehicle trajectories and videos are limited in their coverage as they gather specific routes and/or are captured at ground or close to ground level. OSM has the attraction of openly available data, but at this point, coverage and data quality problems (e.g. accuracy and topological errors) are not fully resolved. These shortcomings make using images the best option.

Images taken from different perspectives have been used in intersection detection, e.g. images from vehicles [19], images from pedestrian perspectives [20], static street level (Google Street View images (GSV)) [11], and remote sensing (overhead distance) images [8], [9], [10]. Images from vehicles and pedestrian perspectives are used to detect intersections in real-time, during a journey, rather than advance route planning. Street level images such as GSV images do not capture the details of big intersections, as street level image coverage is limited. This work focuses on remote-sensing images, which capture roads from sufficient distance for the coverage needed to identify intersections. In previous work, remote-sensing datasets have been limited in size and none of them include the degree of intersection (number of branches) as a label [9], [10], [5]. The size limitation prevents from training deep learning approaches. This was the motivation for constructing the benchmark dataset presented here.

Two high level task categories are identified, in which intersection detection from images is discussed (1) real-time and (2) static. For real-time tasks, intersection shape is detected on-the-fly to help in navigation applications such as autonomous vehicles [11], [21], [19], and robotic navigation [20]. In the static task scenario, the location and shape of intersections are determined to generate persistent knowledge for other tasks [7], [8]. Static tasks are offline processes for extracting map information (such as the location of intersections, traffic lights, public areas, etc.) and road networks. It is considered the backbone of a route selection task [9], [10], which is of particular interest in intersection detection.

Physically, intersection regions can be complicated, frequently obscured by obstructions and shadows of other objects such as buildings, trees, and vehicles [9], [8]. To handle this complexity, deep learning is used. As a state-of-the-art approach in the field of computer vision, deep learning has been applied in many remote-sensing applications such as a road extraction [22], scene classification [23], and automatic building extraction [24]. The problem of intersection detection has been tackled to date using three approaches: classification, road detection and object detection. Kumar et al. [12] detect the existence of intersections in a video using a network consisting of Convolutional Neural Networks (CNN), bi-long short-term memory (LSTM), and Siamese-CNN. Different CNN and LSTM architectures were used to classify images and videos to multiple types of intersections [25], [19], [21]. These previous works did not use satellite images as a source of data and were focused on real-time intersection detection, limiting their benefits towards route planning.

Rebai et al. [13], Costea et al.[8], and Tümen and Ergen [11] used various approaches to detect roads. Then, after detecting roads, Rebai et al. [13] and Tümen and Ergen [11] classified images from vehicles to intersection types, while Costea et al.[8] utilized generative adversarial networks (GANs) to detect intersections from satellite images. In this approach, the accuracy of intersection detection relies on the preceding road extraction stage. Although Costea et al.[8] used satellite images, they did not detect the features of intersection or discuss the partial covering of intersection on satellite images.

Saeedimoghaddam and Stepinski [26] detected intersections from historical map images using Faster RCNNs. In this approach, the image were not up to date, the features of intersection were not detected, and the problem of partial coverage of intersection in images was not addressed. In the approach presented here, deep learning is also used, given its ability to address complex visual tasks such as image classification, object detection and segmentation.

Detecting the locations and features of intersections, and making this static information available is important for applications such as route planning. Looking at research work on this problem, none of the datasets available include both the location of intersection and the degree (number of branches). This limits the use of the intersection information for downstream applications that need to understand intersection features for decision making. Also, most of the previous works treat the problem solely as a classification problem, rather than detecting the location and scale of intersections with object detection. To the best of the authors' knowledge, the problem of partial coverage of intersection in satellite images is not discussed in previous work.

III. Dataset Construction

This dataset aims to capture satellite images that support the detection of road intersections and intersection features. Images cover intersections with different shapes and orientations. Each image is annotated with an intersection class (a positive class) or a no-intersection class (a negative class).

A. *Dataset Capturing*

The aim here is to capture a dataset for an urban city area, using a public aerial image source, to support supervised training of models that will detect the existence of road intersections in aerial images. This dataset uses Google Maps as a source of images. Google Map has two advantages. First, images can be downloaded with a zoom level that captures the details of an intersection area. Second, Google Maps has a global coverage, so the approach used here can be expanded or re-used for other areas. Some decisions need to be made before downloading images, which will be explained here in detail.

Google provides a choice of magnification levels. Zoom level 19 was used here as intersection details are visible at this level. Four types of images can be downloaded using the maps static API: roadmap, terrain, satellite and hybrid [27]. The first two types are maps that are generated by Google. To make this work more general and widely applicable, raw images are preferred as these represent images that could in theory come from any equivalent satellite source. Hybrid images are notated satellites images. They have been labelled by Google to highlight roads and display street names. The satellite images are raw, so roads and intersections may be obscured or hidden by other objects such as trees. The hybrid data has more visual clues but the satellite images are more typical of other data sources. Both were used to create two datasets: a satellite intersection dataset and a hybrid intersection dataset, see Fig. 1.

(a) (b)

Fig. 1 Hybrid and satellite Google Map images for the same location: (a) Satellite image and (b) Hybrid image

The main objective here is to build a model to discover intersections from images in one city. To capture a city at zoom level = 19, tiles that cover the whole city were downloaded. The size of each tile is 256 * 256 pixels (58.98 km^2 geographic size). In this case study, Washington DC is used. It represents a large U.S urban area. It also has intersection location information available via a published Government dataset to assist in ground truth verification. Before using tiles as a source of images and the Government dataset as a labeling source, a quality check was performed on the data sources used.

A quality report was made for tiles that contain intersections. In this quality report, various factors were evaluated; the correctness of the intersection's center and the perfectness of an image's coverage of an intersection. This quality check revealed a major problem with the tile coverage. 76 % of tiles that contain an intersection according to the Government dataset source did not fully cover the entire intersection (i.e. contained a partial intersections). This problem means that many of the images presented partial, incomplete visuals of a road intersection, making their identification more complicated, as shown in Fig. 2. To address this, a way to take a wider view of the tiles is needed, taking bordering tiles into consideration, rather than treating each tile as an isolated artefact. To solve this, a method for capturing an area between tiles was developed. A 256 * 256 window was shifted with 50 % in horizontal and vertical axes to cover each region between tiles, see Fig. 3. Blue rectangles represent tiles. After applying the overlapping method on the first tile, eight new images were created. These new images cover the spaces between the tiles. This method will guarantee that each road intersection will be fully contained in at least one image. This means that if an intersection is partially contained in an image, it will be treated as a negative instance. Using the overlapping method, each partial intersection will be also fully covered in another image. This is based on the assumption that intersections are less than half the width and height of a tile (i.e. 58.98 km^2). Consequently, the model can detect all intersections in a test region. This work resolves the partial coverage of an object in a remote sensing image. The zoom

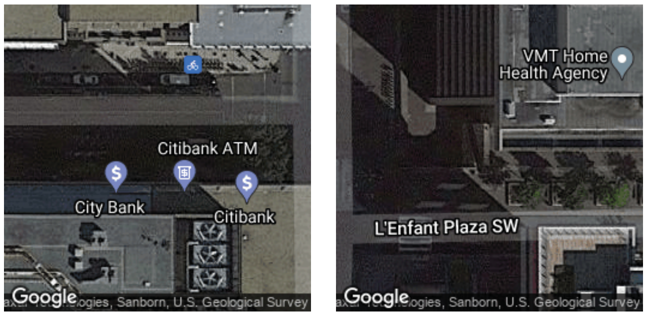


Fig. 2 Partial Intersection Examples

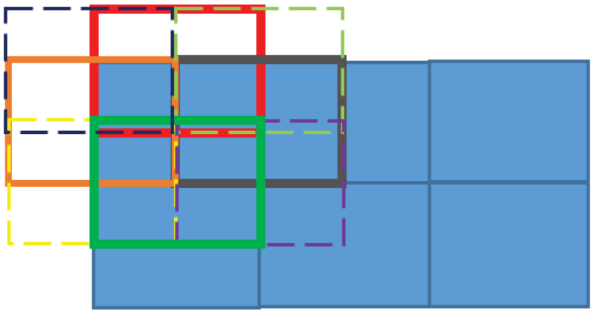


Fig. 3 Overlapping method

level and amount of shifting are suitable for an intersection but they will need to be adjusted again according to an object under consideration.

In this dataset, the image size is 256 * 256 pixels. If the intersection centre is located at least 50 pixels away from the border, it will be considered as an intersection. If it is located in the image within 50 pixels' width around the image, this means that the image partially covers the intersection. It will not be annotated as an intersection. Using the proposed method, the partially covered intersection will be captured and annotated in another image.

B. Dataset Annotation

The first version of the dataset is constructed and annotated to support a classification problem: defining whether an image contains a road intersection or not. Each image was labeled with a positive or negative example. A method was developed to annotate images automatically using the verified intersection points dataset from the Washington DC government website [28]. In this dataset, intersections' locations are defined [29]. If an image covers an area that contains at least one intersection, it was defined as a positive instance (an intersection class), see Fig. 4. Otherwise, it was defined as a negative example (a no-intersection class), see Fig. 5. The positive intersection class includes 7,342 images. The negative class includes 7,350 images. Some images include more than one intersection.

It is important to include negative examples (no-intersection class) in the dataset for training, developing, and testing phases. No-intersection class includes both images that do not include any intersection as shown in Fig. 5 and images that include just part of one intersection as shown in Fig. 2.

Fig. 4 Intersection Examples

Fig. 5 No-intersection Examples

The intersections of roads with private alleys are not included in the intersection points dataset [29]. In Google Map images, they look like regular intersections (roads intersections), see Section VI. These paths can be a footpath, a garage entrance, a bike path, a small path between houses. Fig. 6 contains examples of these paths in google street view. In the images dataset, these have not been labelled as intersections.

To summarise, after capturing and annotating the dataset, two types of datasets were constructed: a satellite intersection dataset and a hybrid intersection dataset. Each one has two classes: intersection and no-intersection. In each dataset, the intersection category contains 7342 images and the no-intersection category contains 7350 images.

C. Dataset Availability

An ongoing problem in this domain is the lack of shared datasets for benchmarking techniques and avoidance of re-labelling. A goal of this work is to make the two datasets constructed here easily available to the research community. Google's terms and conditions prevent direct publication or distribution of the images. In order to make the datasets accessible, an automated script has been provided to generate replica labelled datasets via the Google Maps static API (but using the API key of the downloader). The script will automatically download the exact satellite and hybrid images of our dataset. The script will also automatically apply the same image ground truth annotations as per our datasets. This method does not collide with the Google's use terms and conditions because it does not publish any of Google data. The images used in these datasets were downloaded without cost, using the free allocation of images for Google Maps static API, as of publication date. The automated script and instructions are available to download [30].

IV. Convolutional Neural Networks for Intersection Classification

It is useful to determine how accurately the presence or absence of road intersections in a map tile (i.e. image) can be distinguished using a supervised machine learning approach. Given the cost of acquiring and annotating labelled map data, it is also helpful to investigate how much trained data are really needed to get reasonable results. Specifically, the drop in detection accuracy is examined as the size of training data is gradually reduced. As part of this, models are created both from scratch and using a pre-trained transfer learning approach. The impact on accuracy of each data source (raw satellite and hybrid datasets) is observed. In addition, a comparison between architectures' results in intersection classification is done.

To address an intersection classification problem, a variety of state-of-the-art CNN architectures were trained and evaluated. Xception [31], InceptionResNetV2 [32], and ResNet152V2 [33] were used. These were trained using a satellite intersection dataset and a hybrid intersection dataset (the details about these datasets are available in Section III).

A. Approach and Method

For all experiments, Keras with TensorFlow as a back-end was used. For in-built data augmentation during training, images were flipped horizontally and vertically. At a future stage more elaborate augmentation methods such as shifting and brightness may be considered. In the pre-processing stage, pixel values were normalized. Xception [31], InceptionResNetV2 [32], and ResNet152V2 [33] were used. It is useful to understand the extent to which pre-existing models can be built on, so as to reduce reliance on labelled data and expensive annotation. Therefore, these architectures were trained twice. First, they were trained from scratch. Second, they were fine-tuned. A batch size of 16 was used with the Adam optimizer [34]. The datasets are randomly split to 70 % for training, 20 % for validation, and 10% for testing. To evaluate the classification models, class accuracy (recall) and overall accuracy were used, as shown in Equations 1 and 2.

$$Recall = \frac{\text{True Positives}}{\text{True Positives + False Negatives}} \tag{1}$$

$$Accuracy = \frac{\text{True Positives}}{\text{Total number of images}} \tag{2}$$

In previous work, different metrics were used for evaluation such as accuracy, F1 and confusion matrices. Bhatt et al. [25] identified road intersections from monocular cameras and used accuracy, precision, recall and F1 to measure success. Kumar et al. [12] determined the existence or absence of an intersection in a video and reported accuracy and F1. Oeljeklaus et al. [19] and Koji and Kanji [21] used confusion matrices to evaluated intersection classification. Accuracy

Fig. 6 Google Street View images for a small path intersection

TABLE I
RESULT OF A HYBRID DATASET CLASSIFICATION (TRAINING FROM SCRATCH)

Architectures	Intersection recall	Nointersection recall	Accuracy
Xception	92%	92%	92%
InceptionResNetV2	89%	87%	88%
ResNet152V2	91%	90%	90%

TABLE II
RESULT OF A HYBRID DATASET CLASSIFICATION (TRANSFER LEARNING)

Architectures	Intersection recall	Nointersection recall	Accuracy
Xception	93%	92%	93%
InceptionResNetV2	93%	93%	93%
ResNet152V2	94%	93%	94%

alone does not give a sense of the model's result. F1 is a function of precision and recall. It is used when the model seeks a balance between precision and recall. Recall is more important in the intersection detection problem as there is a high cost associated with false-negatives (image is classified as an intersection while it is not an intersection). A confusion matrix is used to compute the initial calculation about a model but is hard to compare between two models. By using recall, an insight is gained into whether the model is doing the same on both classes – or whether it is stronger at detecting one class over the other.

V. QUANTITATIVE RESULTS

Several experiments were designed to evaluate the ability of deep learning models to classify satellite and hybrid images to intersection and no-intersection classes. Each of the two constructed datasets (satellite and hybrid) were used to train state-of-the-art ConvNet architectures: Xception, InceptionResNetV2, and ResNet152V2. Both datasets are used to train CNNs from scratch and using transfer learning. After training the various CNNs from scratch, the result of detecting the existence of intersection or not in an images is evaluated on test sets, see Tables I and II. After using transfer learning, the results of models is shown in Tables III and IV.

Using a hybrid intersection dataset as training data, Xception model reaches the highest accuracy at 92%, with equal recall on both classes. Using transfer learning, the highest accuracy is achieved using Resnet at 94%. Using a satellite intersection dataset as training data from scratch, Xception is still highest, reaching 89% accuracy in a test set. By using transfer learning it jump up to 90%.

In general, models that were trained on a satellite intersection dataset perform worse than the same models when they were trained on a hybrid intersection dataset. It is observed that visual interruptions such as shadows covering the area of intersection, see Section VI, negatively impact the performance of the satellite model. In addition, hybrid images contain notated visual clues in the vicinity of intersections such as street names, which may guide the model. There is no room to conduct comparative experiments as no similar dataset is available, as shown in Section II.

Labeled datasets are expensive and time consuming to produce. In the following, the proportion of the dataset needed to build a mode which can classify images will be examined. Fig. 7 shows the accuracy of models which are fine-tuned with different percentages of the satellite training dataset. All of them are tested in the same test set. Looking at the slope of the accuracy curve, it can be seen that at 20% of the dataset (2,644 images for both classes), the accuracy is beginning to level off, reaching an accuracy of 77% or less.

VI. QUALITATIVE RESULTS

In this section, examples that were correctly categorized and examples that were incorrectly categorized were studied to understand the model and interpret the cause of errors. Table IV declares that the fine-tuned Xception model achieves the best result on the satellite test set, so it was used for investigation on the result. Fig. 8 shows the qualitative results of the model. Images in Row 1 and Row 2 display that

TABLE III
RESULT OF A SATELLITE DATASET CLASSIFICATION (TRAINING FROM SCRATCH)

Architectures	Intersection recall	Nointersection recall	Accuracy
Xception	90%	89%	89%
InceptionResNetV2	86%	87%	86%
ResNet152V2	86 %	85%	86%

TABLE IV
RESULT OF A SATELLITE DATASET CLASSIFICATION (TRANSFER LEARNING)

Architectures	Intersection recall	Nointersection recall	Accuracy
Xception	91%	89%	90%
InceptionResNetV2	88%	90%	89%
ResNet152V2	88%	85%	87%

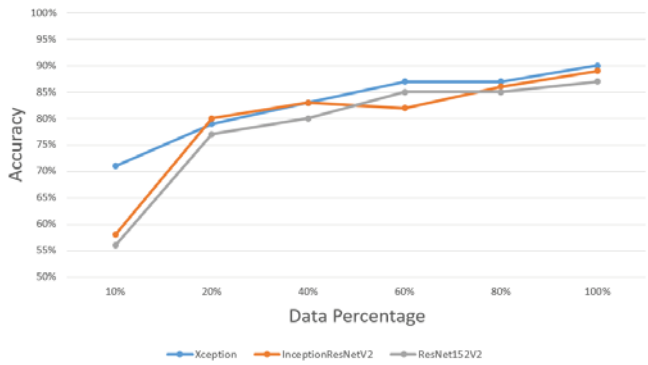


Fig. 7 Accuracies obtained with different fine-tuned models using different percentages of usage satellite dataset: Y-axis represents the accuracy of models while X-axis represents the percentage of used dataset. Blue, orange and gray lines represent the accuracy of Xception, InceptionResNetV2, and ResNet152V2 models

the model can predict a class of an image even if the context is complex (different types, orientations, and sizes of intersections, shadows overlapping). The model incorrectly classifies some images as positive examples for different reasons such as intersection with alleys, bridges overlapping, buildings seem like streets, part of a big intersection (4 degrees) looks like a small intersection (3 degrees), as shown in the Row 3. This raises an interesting issue around the definition of intersection. Our working assumption is that an intersection of roads is at a single level. The false positive examples in Row 3 show roads that cross over (rather than physically intersect) at different heights. They are not considered as intersections in our definition. One way to address this is to increase the number of examples of these types of multi-level intersections as negative examples in the training set, to bridge the semantic gap. On the other hand, the model incorrectly classifies some images as negative examples while they include intersections. The main reasons are tress and trees' shadows which cover part or whole of an intersection, as shown in the last Row. Even for a human being, it is hard to define the intersection if it is fully covered by trees.

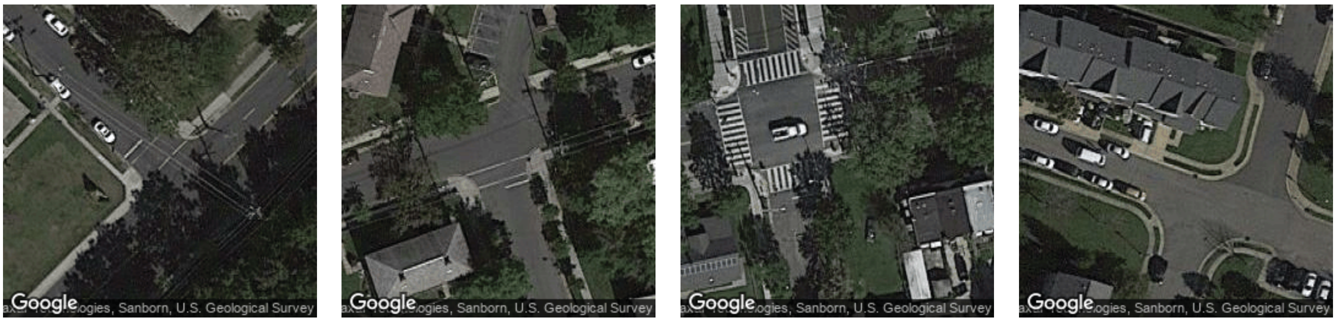
VII. CONCLUSION AND FUTURE WORK

Pre-defining road intersections is a significant task for applications such as routing planning and road network extraction. This paper presented a substantial dataset for road intersection detection. Two types of data were gathered: a raw satellite images dataset and a hybrid dataset with street name notations. Each dataset includes 14,692 images from Washington DC. The construction and annotation of the datasets are explained, including an automated tool for labelled dataset replication by other researchers. A overlapping sliding window approach is presented to address the issue of objects split across multiple satellite images. Three state-of-the-art CNN architectures were utilized to determine how accurately the existence of intersections can be detected from satellite images. The results of models that are trained from scratch or fine-tuned using both datasets were compared. The fine-tuned ResNet152V2 model trained on hybrid images achieved the best accuracy of 94%. About 40% of each dataset (5,289 images) can be used to train models and reach an accuracy of 80% or greater.

In the next stage of the work, it is planned to extend the labeling of the datasets to include the location and number of branches in each intersection. Object detection models will also be used to detect the intersections and define their degree in images.

ACKNOWLEDGMENT

This publication has emanated from research supported in part by a Grant from Science Foundation Ireland under Grant number 18/CRT/6222. It has also funded by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Co-funding of regional, national and international programmes (Grant agreement No. 713654)



(a)

(b)

(c)

(d)

Fig. 8 Qualitative results: (a) True positive examples, (b) True negative examples, (c) False positive examples, and (d) False negative examples

REFERENCES

[1] Q. Zhou and Z. Li, "Experimental analysis of various types of road intersections for interchange detection," *Transactions in GIS*, vol. 19, no. 1, pp. 19–41, 2015.

[2] P. Fogliaroni, D. Bucher, N. Jankovic, and I. Giannopoulos, "Intersections of our world," in *Proceedings of 10th international conference on geographic information science*, vol. 114. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018, p. 3.

[3] S. Fang, L. Yang, T. Wang, and S. Jing, "Trajectory planning method for mixed vehicles considering traffic stability and fuel consumption at the signalized intersection," *Journal of Advanced Transportation*, vol. 2020, 2020.

[4] B. Asadi and A. Vahidi, "Predictive cruise control: Utilizing upcoming traffic signal information for improving fuel economy and reducing trip

time," *IEEE transactions on control systems technology*, vol. 19, no. 3, pp. 707–714, 2010.

[5] F. E.-Z. El-Taher, A. Taha, J. Courtney, and S. Mckeever, "A systematic review of urban navigation systems for visually impaired people," *Sensors*, vol. 21, no. 9, p. 3103, 2021.

[6] A. Cohen and S. Dalyot, "Route planning for blind pedestrians using openstreetmap," *Environment and Planning B: Urban Analytics and City Science*, vol. 48, no. 6, pp. 1511–1526, 2021.

[7] P. Li, Y. Li, J. Feng, Z. Ma, and X. Li, "Automatic detection and recognition of road intersections for road extraction from imagery," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43, pp. 113–117, 2020.

[8] D. Costea, A. Marcu, E. Slusanschi, and M. Leordeanu, "Creating roadmaps in aerial images with generative adversarial networks and smoothing-based optimization," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 2100–2109.

[9] J. Dai, Y. Wang, W. Li, and Y. Zuo, "Automatic method for extraction of complex road intersection points from high-resolution remote sensing images based on fuzzy inference," *IEEE Access*, vol. 8, pp. 39 212–39 224, 2020.

[10] J. Cheng, T. Liu, Y. Zhou, and Y. Xiong, "Road junction identification in high resolution urban sar images based on svm," in *Proceedings of the International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*. Springer, 2019, pp. 597–606.

[11] V. Tümen and B. Ergen, "Intersections and crosswalk detection using deep learning and image processing techniques," *Physica A: Statistical Mechanics and its Applications*, vol. 543, p. 123510, 2020.

[12] A. Kumar, G. Gupta, A. Sharma, and K. M. Krishna, "Towards view-invariant intersection recognition from videos using deep network ensembles," in *Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1053–1060.

[13] K. Rebai, N. Achour, and O. Azouaoui, "Road intersection detection and classification using hierarchical svm classifier," *Advanced Robotics*, vol. 28, no. 14, pp. 929–941, 2014.

[14] S. Kumaar, S. Mannar, S. Omkar *et al.*, "Juncnet: A deep neural network for road junction disambiguation for autonomous vehicles," *arXiv preprint arXiv:1809.01011*, 2018.

[15] U. Baumann, Y.-Y. Huang, C. Gläser, M. Herman, H. Banzhaf, and J. M. Zöllner, "Classifying road intersections using transfer-learning on a deep neural network," in *Proceedings of the International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 683–690.

[16] J. Bock, R. Krajewski, T. Moers, S. Runde, L. Vater, and L. Eckstein, "The ind dataset: A drone dataset of naturalistic road user trajectories at german intersections," in *Proceedings of 2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2020, pp. 1929–1934.

[17] J. Wang, C. Wang, X. Song, and V. Raghavan, "Automatic intersection and traffic rule detection by mining motor-vehicle gps trajectories," *Computers, Environment and Urban Systems*, vol. 64, pp. 19–29, 2017.

[18] J. HUANG, M. DENG, Y. ZHANG, and H. LIU, "Complex road intersection modelling based on low-frequency gps track data," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, no. 2/W7, 2017.

[19] M. Oeljeklaus, F. Hoffmann, and T. Bertram, "A combined recognition and segmentation model for urban traffic scene understanding," in *Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2017, pp. 1–6.

[20] M. Astrid, J.-H. Lee, M. Z. Zaheer, J.-Y. Lee, and S.-I. Lee, "For safer navigation: Pedestrian-view intersection classification," in *Proceedings of the 2020 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2020, pp. 7–10.

[21] T. Koji and T. Kanji, "Deep intersection classification using first and third person views," in *Proceedings of the 2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2019, pp. 454–459.

[22] A. Abdollahi, B. Pradhan, and A. Alamri, "Vnet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data," *IEEE Access*, vol. 8, pp. 179 424–179 436, 2020.

[23] J. Zhang, C. Lu, X. Li, H.-J. Kim, and J. Wang, "A full convolutional network based on densenet for remote sensing scene classification," *Mathematical Biosciences and Engineering*, vol. 16, no. 5, pp. 3345–3367, 2019.

[24] Z. Shao, P. Tang, Z. Wang, N. Saleem, S. Yam, and C. Sommai, "Brrnet: A fully convolutional neural network for automatic building extraction from high-resolution remote sensing images," *Remote Sensing*, vol. 12, no. 6, 2020.

[25] D. Bhatt, D. Sodhi, A. Pal, V. Balasubramanian, and M. Krishna, "Have i reached the intersection: A deep learning-based approach for intersection detection from monocular cameras," in *Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 4495–4500.

[26] M. Saeedimoghaddam and T. Stepinski, "Automatic extraction of road intersection points from usgs historical map series using deep convolutional neural networks," *International Journal of Geographical Information Science*, pp. 1–22, 2019.

[27] "Google map types," https://developers.google.com/maps/documentation/maps-static/start#MapTypes (archived on 2021-7-26).

[28] "Government website of washington dc datasets," https://opendata.dc.gov/ (archived on 2021-7-26).

[29] "Intersection points dataset," https://opendata.dc.gov/datasets/DCGIS::intersection-points/about (archived on 2021-7-26).

[30] "Script to download road intersections dataset," https://github.com/fatmaelther/SatelliteDatasetsforRoadIntersectionDetection (archived on 2021-12-10).

[31] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[32] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the Thirty-first AAAI conference on artificial intelligence*, 2017.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proceedings of European conference on computer vision*. Springer, 2016, pp. 630–645.

[34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.